

# Variance-reduced multiscale simulation of slow-fast stochastic differential equations

Ward Melis<sup>a,\*</sup>, Giovanni Samaey<sup>a</sup>

<sup>a</sup>*Department of Computer Science, K.U. Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium*

---

## Abstract

We study a variance reduction strategy based on control variables for simulating the averaged macroscopic behavior of a stochastic slow-fast system. We assume that this averaged behavior can be written in terms of a few slow degrees of freedom, and that the fast dynamics is ergodic for every fixed value of the slow variable. The time derivative for the averaged dynamics can then be approximated by a Markov chain Monte Carlo method. The variance-reduced scheme that is introduced here uses the previous time instant as a control variable. We analyze the variance and bias of the proposed estimator and illustrate its performance when applied to a linear and nonlinear model problem.

*Keywords:* variance reduction, slow-fast stochastic systems, multiscale simulation

---

## 1. Introduction

Stochastic differential equations (SDEs) are ubiquitous in a multitude of real-life applications, appearing in different scientific domains such as climate and environmental sciences [1, 2, 3], molecular dynamics [4, 5] and bacterial chemotaxis [6]. Many of these applications contain processes that inherently evolve over multiple time scales, leading to excessive computational cost with standard time discretization methods. As a consequence, there is currently a large interest in developing dedicated numerical methods that circumvent, or even exploit, the presence of a time-scale separation in the problem at hand.

Clearly, the development of new numerical techniques needs to be supplemented by a detailed analysis of their efficiency and accuracy, and this for a set of model problems of which the multiscale nature is well understood. One prototypical example system that was proposed in [5] to analyze such convergence behavior is a singularly perturbed slow-fast system in which the slow variable is described deterministically, while the model for the fast variable contains stochastic effects. The specific form is as follows:

$$\begin{cases} dx(t) = f(x, y)dt, & x(0) = x_0 \in \mathbb{R} \\ dy(t) = \frac{1}{\varepsilon}g(x, y)dt + \frac{1}{\sqrt{\varepsilon}}\beta(x, y)dW(t), & y(0) = y_0 \in \mathbb{R}. \end{cases} \quad (1)$$

where the scalar quantities  $x(t) : [0, T] \rightarrow \mathbb{R}$  and  $y(t) : [0, T] \rightarrow \mathbb{R}$  represent the slow and fast evolving stochastic processes, respectively. The functions  $f(x, y), g(x, y) \in \mathbb{R}$  are called the drift functions and  $\beta(x, y) \in \mathbb{R}$  is termed the diffusion function. Furthermore,  $W(t) \in \mathbb{R}$  denotes a standard Brownian motion. The parameter  $\varepsilon \ll 1$  is a positive small-scale parameter that measures the time scale separation between the fast and slow variable in system (1). In addition, we assume that the fast dynamics is ergodic for every fixed state  $X \in \mathbb{R}$  of the slow variable, implying the existence and uniqueness of an invariant measure [7]. We note that the differential form used in system (1) is purely formal, given that Brownian paths are continuous everywhere but nowhere differentiable. Consequently, system (1) should be understood in the

---

\*Corresponding author

Email addresses: ward.melis@cs.kuleuven.be (Ward Melis), giovanni.samaey@cs.kuleuven.be (Giovanni Samaey)

integral form, where stochastic integrals are interpreted in the Itô-sense. In general, the SDE may be very high-dimensional (especially with many fast degrees of freedom), see, for instance, [5].

Often, one is only interested in the evolution of the slow variable of system (1) and not in the detailed evolution of the fast variable. However, the fast dynamics cannot be omitted, since the slow process explicitly depends on the fast variable. Due to the stiffness in system (1), explicit simulation techniques such as the Euler-Maruyama or higher-order Milstein schemes are computationally prohibitive. Also, implicit methods fail to capture the correct invariant measure, thus introducing a bias, see [8].

The difficulties related to direct simulation can be avoided by exploiting the time scale separation of system (1): for  $\varepsilon \rightarrow 0$ , the averaging principle yields a reduced description for the slow variable:

$$\frac{dX}{dt} = F(X), \quad F(X) = \int_{\mathcal{Y}} f(X, y) d\mu_X^\infty(y), \quad (2)$$

in which  $\mu_X^\infty(y)$  denotes the invariant measure induced by the fast dynamics of system (1) keeping  $x = X$  fixed, see, for instance, [7] and references therein. Equation (2) is known as the *averaged, macroscopic* or *reduced* equation for the slow variable.

Based on this averaged equation, a method for (1) was proposed in [9] and analyzed in [10]. It consists of a macroscopic solver, such as the forward Euler or a higher-order Runge-Kutta method, to simulate (2), combined with a procedure to estimate the effective force  $F(X)$  in equation (2). If the invariant measure is known explicitly and can readily be sampled, the integral in (2) can be approximated by a direct Monte Carlo estimator (see, e.g., [11] and references therein). In general, however, the invariant measure is not known explicitly. Then, one may resort to a Markov chain Monte Carlo method, as is done in [9]. This method fits in the class of *heterogeneous multiscale methods* (HMM) that were introduced in [12] for a broad class of multiscale problems and provide a natural setting for numerical analysis, see also [13] for a recent review. Similar methods have been introduced based on the concept of *coarse projective integration* [14]. There, instead of performing one (or a few) long Markov chain Monte Carlo simulation of the fast equation, one initializes a *large ensemble* of realizations, which are simulated on a short time interval. This method falls in the class of *equation-free methods* [15, 16], see also [17], and can also be used in a more general setting where one is unable to identify or constrain the slow degree of freedom.

Unfortunately, the statistical error of the above-described methods can be quite large, and decreases only as  $M^{-1/2}$  when the number of samples  $M$  tends to infinity. In this work, which expands the results reported in [18], we therefore propose a variance-reduction technique based on control variables, see, e.g., [11, 19]. The method can be applied both in the projective integration and the HMM setting, and bears some resemblance to the technique that was proposed in [20] for variance-reduced coarse projective integration of SDEs of the form (1). While we present the method and main analysis in the HMM framework, we will comment on coarse projective integration where appropriate. The control variable that we introduce is based on correlating estimations of the time derivative  $F(X)$  in equation (2) on different time instants.

The remainder of this paper is structured as follows. In section 2, we introduce the stochastic slow-fast system that we intend to solve numerically. In section 3, we describe the HMM framework to efficiently integrate these slow-fast systems. In that section, we also present the variance-reduced HMM method, and we comment on the applicability of this method for coarse projective integration. Next, in section 4 we analyze the numerical properties of the proposed variance reduction method. Numerical results are reported in section 5. We conclude in section 6 with a brief discussion and ideas for future work.

## 2. Slow-fast system

The general form of slow-fast systems we consider in this work is given in equation (1). In what follows, we will always assume that the fast dynamics of system (1) is ergodic for all fixed values of the slow variable. This means that the fast equation produces a unique invariant measure for every fixed value  $X$ , denoted by  $\mu_X^\infty(y)$ . Ergodicity implies that the statistical properties of the ensemble of the stochastic process at a fixed time instant and those of one realization of the process over an infinite time interval are the same.

Consequently, for an ergodic process, averaging a function with respect to the invariant measure yields the same result as averaging this function over one infinitely long time path of the process:

$$F(X) = \int_{\mathcal{Y}} f(X, y) d\mu_X^\infty(y) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(X, y(t + \tau)) d\tau. \quad (3)$$

Equation (3) serves as a base for numerical methods avoiding explicit knowledge of the invariant measure  $\mu_X^\infty(y)$ . Additionally, we will always assume that the invariant measure possesses a density  $\rho_X^\infty$  with respect to the Lebesgue measure:  $d\mu_X^\infty(y) = \rho_X^\infty(y) dy$ .

Moreover, we assume that the function  $F(X)$  in equation (2) is Lipschitz continuous with Lipschitz constant  $L_c$ , implying the following inequality:

$$|F(X_1) - F(X_2)| \leq L_c |X_1 - X_2|, \quad \forall X_1, X_2 \in \mathbb{R}, \quad (4)$$

and we assume the functions  $f$ ,  $g$  and  $\beta$  sufficiently differentiable such that all derivatives exist that are required during the analysis.

In the following two paragraphs, we introduce the linear and nonlinear illustrative examples that will be used in the numerical experiments throughout the text.

*Linear system.* In the linear setting, system (1) takes on the following form:

$$\begin{cases} dx(t) = (\lambda x(t) + py(t)) dt \\ dy(t) = \frac{1}{\varepsilon} (qx(t) - Ay(t)) dt + \frac{1}{\sqrt{\varepsilon}} dW(t), \end{cases} \quad (5)$$

in which the parameters  $\lambda$ ,  $p$ ,  $q$  and  $A$  are all real scalars. In addition, to ensure that solutions decay exponentially with time, we require that  $\lambda < 0$  and  $A \in \left( \frac{pq}{-\lambda}, 2 \right]$ .

For this linear system, the fast equation corresponds to a linear Ornstein-Uhlenbeck process with parameters  $q$  and  $A$  for which the invariant measure can be calculated analytically as [7]:

$$\mu_X^\infty(y) \sim \mathcal{N}(m_\infty, \sigma_\infty^2), \quad m_\infty = \frac{q}{A} X, \quad \sigma_\infty^2 = \frac{1}{2A}, \quad (6)$$

where  $\mathcal{N}(\cdot, \cdot)$  represents the normal distribution and  $m_\infty$  and  $\sigma_\infty^2$  denote the mean and variance of the invariant measure, respectively. Since the invariant measure is known, the integral in equation (2) can be calculated analytically, yielding:

$$F(X) = \int_{\mathcal{Y}} (\lambda X + py) d\mu_X^\infty(y) = \lambda X + p \int_{\mathcal{Y}} y d\mu_X^\infty(y) = \left( \lambda + \frac{pq}{A} \right) X. \quad (7)$$

In that case, the reduced equation (2) for system (5) becomes:

$$\frac{dX}{dt} = \left( \lambda + \frac{pq}{A} \right) X, \quad (8)$$

which is a linear ODE. The exact solution to equation (8) with initial condition  $X(0) = x_0$  is then given as:

$$X(t) = x_0 \exp\left(\left(\lambda + \frac{pq}{A}\right)t\right). \quad (9)$$

*Nonlinear system.* As a second example, we consider the following nonlinear stochastic multiscale system from [20]:

$$\begin{cases} dx(t) = -(y(t) + y(t)^2) dt \\ dy(t) = -\frac{1}{\varepsilon} (y(t) - x(t)) dt + \frac{1}{\sqrt{\varepsilon}} dW(t). \end{cases} \quad (10)$$

In this case, the dynamics of the slow variable is nonlinear, while the fast variable is again described by a linear Ornstein-Uhlenbeck process. Using equation (6), we obtain a Gaussian invariant measure  $\mu_X^\infty(y)$  with the invariant mean  $m_\infty = X$  and variance  $\sigma_\infty^2 = 1/2$ . Since the invariant measure is known explicitly, the expression of  $F$  in equation (2) can be calculated analytically as:

$$F(X) = - \int_{\mathcal{Y}} y d\mu_X^\infty(y) - \int_{\mathcal{Y}} y^2 d\mu_X^\infty(y) = - \left( X + X^2 + \frac{1}{2} \right), \quad (11)$$

resulting in a nonlinear ODE. The exact solution of the resulting macroscopic equation with initial condition  $X(0) = x_0$  is obtained as:

$$X(t) = -\frac{1}{2} - \frac{1}{2} \tan \left( \frac{t}{2} - \arctan(2x_0 + 1) \right). \quad (12)$$

### 3. Numerical method

In this section, we construct a variance-reduced numerical scheme to solve the averaged equation (2) for the slow variable of the underlying slow-fast system given in (1). Since the averaged equation is deterministic, any stable explicit ODE solver can be used. Here, we employ the forward Euler (FE) method. To that end, we discretize equation (2) on a uniform time mesh with time step  $\Delta t$ , and  $t^n = n\Delta t$ . The numerical solution on this mesh is denoted by  $\hat{X}^n$ . The forward Euler scheme for (2) is then given by,

$$\hat{X}^{n+1} = \hat{X}^n + \Delta t \hat{F}(\hat{X}^n), \quad \hat{X}^0 = x_0. \quad (13)$$

In equation (13), the function  $F$  is replaced by an appropriate estimator  $\hat{F}$ , since, in general, the integral in equation (2) can not be calculated analytically.

We first introduce the HMM estimator in section 3.1, where we also briefly comment on its relation to coarse projective integration. Then, we present the variance-reduced HMM estimator, which forms the focus of this paper, in section 3.2.

#### 3.1. Heterogeneous multiscale method (HMM)

The heterogeneous multiscale method [9] bypasses explicit knowledge of the invariant measure in equation (2) by exploiting the ergodicity property given in equation (3):  $F$  is calculated by averaging over one infinitely long time path of the fast process of system (1) while keeping the value of the slow variable fixed. As a result, the HMM estimator boils down to a Markov chain Monte Carlo estimator: the integral in equation (2) is approximated by a Monte Carlo method, in which the samples are not drawn from the (unknown) invariant measure, but are instead generated from a Markov chain. This chain is obtained by simulating the fast equation using the explicit Euler-Maruyama scheme, which is the stochastic counterpart of the forward Euler scheme [21]. In that regard, we discretize the fast equation on a uniform time mesh with time step  $\delta t$ . For a given fixed value  $\hat{X}^n$  of the slow variable, the numerical solution at time  $t^{n,m'} = n\Delta t + m'\delta t$  is denoted by  $y^{n,m'}$ . The Euler-Maruyama scheme is given by,

$$y^{n,m'+1} = y^{n,m'} + \frac{\delta t}{\varepsilon} g(\hat{X}^n, y^{n,m'}) + \sqrt{\frac{\delta t}{\varepsilon}} \beta(\hat{X}^n, y^{n,m'}) \xi_n^{m'}, \quad m' = 0, \dots, M-1, \quad (14)$$

in which  $(\xi_n^{m'})_{m'=0}^{M-1}$  is a set of mutually independent samples drawn from the standard normal distribution using a random number generator with seed  $\omega_n$ . The initial condition of the Euler-Maruyama method is chosen as  $y^{0,0} = y_0$ , and for all other  $n > 0$  as  $y^{n,0} = y^{n-1,M-1}$ . Then, the samples generated by the Markov chain (14) are approximately distributed according to the desired invariant measure. To eliminate the time discretization error that the Euler-Maruyama scheme induces in the invariant measure, one could add a Metropolis accept/reject step, as in the MALA algorithm [22].

Since it is more natural to label samples from 1 to  $M$ , we use the trivial substitution  $m = m' + 1$  as sample index. Then, the HMM estimator at time instant  $t^n$  using  $M$  samples is calculated as follows:

$$\hat{F}_M^{\text{HMM}}(\hat{X}^n; \omega_n) = \frac{1}{M} \sum_{m=1}^M f(\hat{X}^n, y^{n,m}), \quad (15)$$

in which  $\omega_n$  represents the seed that is used in the random number generator. Since we can only generate finite sample sizes  $M$ , the HMM estimator in equation (15) is a random variable.

**Remark 3.1 (Coarse projective integration).** The coarse projective integration (CPI) method that was presented in [14] is very similar to the method above, with a different starting point. In [14], one does not assume to be able to simulate the fast equation separately. Instead, to advance from  $\hat{X}^n$  to  $\hat{X}^{n+1}$ , one only performs short-term simulations over a time interval of size, say,  $K\delta t$  with the original system (1), starting from an ensemble of initial conditions  $\left\{ \left( \hat{X}^n, y^{n,m} \right) \right\}_{m=1}^M$ , yielding the time-evolved ensemble  $\left\{ \left( \hat{X}_{K\delta t}^n, y_{K\delta t}^{n,m} \right) \right\}_{m=1}^M$ . The time derivative estimator can then be obtained as

$$\hat{F}_M^{\text{CPI}}(\hat{X}^n) = \frac{1}{M} \sum_{m=1}^M \frac{\hat{X}_{K\delta t}^n - \hat{X}^n}{K\delta t}. \quad (16)$$

While the HMM and CPI methods result in somewhat different equations and have different parameters that can be chosen, the schemes are very similar. In particular, when generating the ensemble of initial conditions for CPI using the Euler-Maruyama method (14) and choosing the number of microscopic time steps  $K = 1$  in the CPI method, both methods can be seen to be identical. The results that are obtained in this paper for the HMM method can therefore easily be carried over to the CPI case.

### 3.2. Variance-reduced HMM

The statistical error on the estimator (15) decays only slowly (as  $M^{-1/2}$ ) with increasing sample size  $M$ , which is typical for any Monte Carlo-based estimator. Moreover, in the HMM method, the samples  $(y^{n,m})_{m=1}^M$  generated by the Markov chain (14) are clearly correlated, resulting in a higher statistical error than that of a Monte Carlo estimator using independent samples. We refer to [23] for an overview on the convergence of Markov chain Monte Carlo sampling. Whenever the statistical error dominates the systematic error, one should reduce the variance. Here, we propose a variance-reduced estimator based on the control variable technique, see, e.g., [11].

*Main idea.* To clearly distinguish between the standard and variance-reduced methods, we will always denote variance-reduced estimates with an overbar, whereas standard estimates will be indicated with a hat. Thus, with the variance-reduced estimator, the forward Euler scheme (13) becomes:

$$\bar{X}^{n+1} = \bar{X}^n + \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^n), \quad \bar{X}^0 = x_0. \quad (17)$$

We define the variance-reduced HMM estimator at time instant  $t^n$  using  $M$  samples as follows:

$$\bar{F}_M^{\text{HMM}}(\bar{X}^n) = \hat{F}_M^{\text{HMM}}(\bar{X}^n; \omega_n) - \left( \hat{F}_M^{\text{HMM}}(\bar{X}^{n-1}; \omega_n) - \bar{F}_M^{\text{HMM}}(\bar{X}^{n-1}) \right), \quad (18)$$

in which an overbar denotes a variance-reduced estimator. The first term  $\hat{F}_M^{\text{HMM}}(\bar{X}^n; \omega_n)$  in equation (18) coincides with the classical HMM estimator without variance reduction for the slow variable  $\bar{X}^n$  at the current time instant using seed  $\omega_n$ . The second term  $\hat{F}_M^{\text{HMM}}(\bar{X}^{n-1}; \omega_n)$  represents another HMM estimation without variance reduction for the slow variable  $\bar{X}^{n-1}$  at the previous time instant. However, this term uses the same seed  $\omega_n$  as the first term and therefore  $\hat{F}_M^{\text{HMM}}(\bar{X}^{n-1}; \omega_n)$  and  $\hat{F}_M^{\text{HMM}}(\bar{X}^n; \omega_n)$  will be strongly correlated. The variance reduction is achieved by subtracting these two terms in an attempt to cancel out the

corresponding statistical variations. The last term  $\bar{F}_M^{\text{HMM}}(\bar{X}^{n-1})$  is the variance-reduced HMM estimator calculated during the previous time step, which needs to be added to avoid introducing a bias.

The proposed technique can also be viewed from the following perspective. The difference between  $\hat{F}_M^{\text{HMM}}(\bar{X}^{n-1}; \omega_n)$  and  $\bar{F}_M^{\text{HMM}}(\bar{X}^{n-1})$  between brackets in equation (18) has zero expectation and approximately corresponds to the noise on the estimator  $\hat{F}_M^{\text{HMM}}(\bar{X}^n; \omega_n)$  since the same seed  $\omega_n$  is used in the first term. The variance reduction method is illustrated in figure 1.

*Initialization.* To get started, the procedure requires a variance-reduced estimation  $\bar{F}_M^{\text{HMM}}(\bar{X}^0)$  in the first step. There are several options:

- *Exact solution.* In some of our numerical experiments, we will choose  $\bar{F}_M^{\text{HMM}}(\bar{X}^0)$  to be the exact solution  $F(\bar{X}^0)$ . Clearly, this is a choice that cannot be made in practical applications (since it is not necessary to use the HMM method when this is possible), so this will only be done to illustrate some properties of the numerical scheme, most notably when studying the bias in section 4.2.
- *More accurate HMM estimator.* A second option is to use a more accurately estimated value, denoted by  $\hat{F}_{M^*}^{\text{HMM}}(\bar{X}^0; \omega_0)$ , with a number of samples  $M^* \gg M$ .
- *Average of HMM estimators.* As a third option, one could also use an average of  $S$  HMM estimators with  $M$  realizations,

$$\hat{F}_{S,M}^{\text{HMM}}(\bar{X}^0; \omega_0) = \frac{1}{S} \sum_{s=1}^S \hat{F}_M^{\text{HMM}}(\bar{X}^0; \omega_{0,s}), \quad (19)$$

where  $\omega_0 = (\omega_{0,s})_{s=1}^S$  represents the vector of initial seeds. Notice that, when choosing  $S = M^*/M$ , computing (19) has the same computational cost as  $\hat{F}_{M^*}^{\text{HMM}}(\bar{X}^0; \omega_0)$ .

*Reinitialization.* Because the invariant measure that is sampled by the Markov chain (14) is parametrized by the slow variable  $X$ , which itself evolves as a function of (macroscopic) time, we expect the variance reduction to become less effective as time advances, see also the analysis in section 4.1. To reduce this effect, we introduce an additional reinitialization step: after every  $R$  macroscopic time steps, we do not compute the variance reduced estimate as in (18), but instead use the initialization procedure described above.

**Remark 3.2 (Coarse projective integration).** The above procedure can also be used for the coarse projective integration estimator (16), provided that the generation of the ensemble of initial conditions  $\{y^{n,m}\}_{m=1}^M$  is done using the Euler-Maruyama scheme (14), and one ensures that the same seed  $\omega_n$  is used when generating these initial conditions as well as for the Brownian increments to compute the time-evolved states  $\{y_{K\delta t}^{n,m}\}_{m=1}^M$  in the two estimations.

## 4. Numerical properties

The convergence of the HMM method described in section 3.1 has been studied in detail in the literature [10, 24], see also [14, 20] for related results. In general, any HMM estimator contains errors from different sources. First, while one intends to exploit the ergodicity property (3), one can only simulate the fast dynamics over a finite time interval  $[0, \tau]$  with  $\tau = M\delta t$ , which leads to a finite sampling error. Second, since the exact solution of the fast equation is not known explicitly, a time discretization method is used to approximate the solution of this equation, which leads to a discretization error. Third, we introduce an error by replacing the (finite) time integral by a finite Riemann sum, in which the fast variable is evaluated at discrete time instants, which leads to a sampling error.

In the present paper, we are not concerned with these errors. We only study the reduction of the variance that results from superimposing the variance reduction technique of section 3.2 onto the HMM estimator (section 4.1). Subsequently, we study the potential *additional* bias of the variance-reduced estimator (18) with respect to the standard HMM estimator (15) in section 4.2.

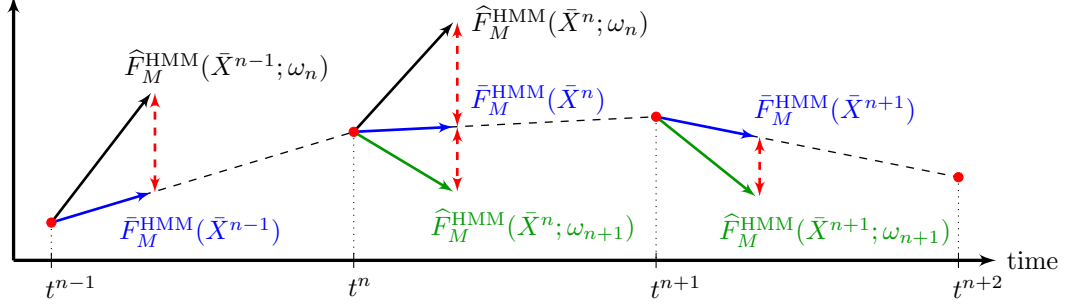


Figure 1: Sketch of the proposed variance reduction technique. Blue arrows depict variance-reduced HMM estimators, which at every time instant require the calculation of two correlated original HMM estimators at the same and previous time instant. These original estimators are shown by black and green arrows for estimation at time  $t^n$  and  $t^{n+1}$ , respectively. The red dashed arrows indicate that, due to the correlation, the estimator noise – the difference between the black (or green) and blue arrows – is similar at consecutive time instants.

#### 4.1. Estimator variance

##### 4.1.1. General case

First, we consider the general (nonlinear) case and study the statistical error, which is quantified by the variance of the estimator:

$$\text{Var}[\bar{F}_M^{\text{HMM}}(\bar{X}^N)] = \text{Var}\left[\hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N) - \left(\hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N) - \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1})\right)\right]. \quad (20)$$

Since the same Brownian path is used twice, we expect  $\hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N)$  and  $\hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N)$  to be strongly correlated and the variance in the estimator  $\bar{F}_M^{\text{HMM}}(\bar{X}^N)$  reduced. To obtain an expression for  $\text{Var}[\bar{F}_M^{\text{HMM}}(\bar{X}^N)]$ , we first rewrite the equation (18). Starting from (18) and using (15), we get:

$$\hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N) - \hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N) = \frac{1}{M} \sum_{m=1}^M \left( f(\bar{X}^N, y_N^{N,m}) - f(\bar{X}^{N-1}, y_N^{N-1,m}) \right), \quad (21)$$

in which  $(y_n^{n,m})_{m=1}^M$  denotes the set of Markov chain generated samples at time  $t^n$  using a random number generator with seed  $\omega_n$ . A Taylor expansion around  $(\bar{X}^N, y_N^{N,m})$  of the difference within the summation of equation (21) leads to:

$$\begin{aligned} f(\bar{X}^N, y_N^{N,m}) - f(\bar{X}^{N-1}, y_N^{N-1,m}) &\approx \partial_x f^m \cdot (\bar{X}^N - \bar{X}^{N-1}) + \partial_y f^m \cdot (y_N^{N,m} - y_N^{N-1,m}) \\ &= \partial_x f^m \cdot \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) + \partial_y f^m \cdot (y_N^{N,m} - y_N^{N-1,m}), \end{aligned} \quad (22)$$

where we used that fact that  $\bar{X}^N$  is obtained using a forward Euler step starting from  $\bar{X}^{N-1}$  (see equation (17)) and introduced the shorthand notation  $\partial_x f^m$  and  $\partial_y f^m$  to denote the partial derivative of the function  $f$  with respect to  $x$  and  $y$ , respectively, evaluated at  $(\bar{X}^N, y_N^{N,m})$  (since, for given  $N$  the argument for which the partial derivative is evaluated is completely determined by  $m$ ). Next, the difference between the samples of the fast equation in the second term of equation (22) can be obtained by subtracting the Markov chains in equation (14) that generate them. For  $m = 1, \dots, M-1$ , this becomes:

$$\begin{aligned} y_N^{N,m+1} - y_N^{N-1,m+1} &= y_N^{N,m} - y_N^{N-1,m} + \frac{\delta t}{\varepsilon} (g(\bar{X}^N, y_N^{N,m}) - g(\bar{X}^{N-1}, y_N^{N-1,m})) \\ &\quad + \sqrt{\frac{\delta t}{\varepsilon}} (\beta(\bar{X}^N, y_N^{N,m}) - \beta(\bar{X}^{N-1}, y_N^{N-1,m})) \xi_N^m. \end{aligned} \quad (23)$$

We again use a Taylor expansion of the functions  $g(x, y)$  and  $\beta(x, y)$  around  $(\bar{X}^N, y_N^{N,m})$  in (23), yielding:

$$y_N^{N,m+1} - y_N^{N-1,m+1} \approx y_N^{N,m} - y_N^{N-1,m} + \frac{\delta t}{\varepsilon} \left( \partial_x g^m \cdot (\bar{X}^N - \bar{X}^{N-1}) + \partial_y g^m \cdot (y_N^{N,m} - y_N^{N-1,m}) \right) + \sqrt{\frac{\delta t}{\varepsilon}} \left( \partial_x \beta^m \cdot (\bar{X}^N - \bar{X}^{N-1}) + \partial_y \beta^m \cdot (y_N^{N,m} - y_N^{N-1,m}) \right) \xi_N^m. \quad (24)$$

Equation (24) can be compactly rewritten as:

$$y_N^{N,m+1} - y_N^{N-1,m+1} \approx A_N^m (y_N^{N,m} - y_N^{N-1,m}) + B_N^m \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}), \quad m = 1, \dots, M-1, \quad (25)$$

in which the random numbers  $A_N^m$  and  $B_N^m$  are given by:

$$A_N^m = 1 + \frac{\delta t}{\varepsilon} \partial_y g^m + \sqrt{\frac{\delta t}{\varepsilon}} \partial_y \beta^m \cdot \xi_N^m, \quad B_N^m = \frac{\delta t}{\varepsilon} \partial_x g^m + \sqrt{\frac{\delta t}{\varepsilon}} \partial_x \beta^m \cdot \xi_N^m. \quad (26)$$

Working out equation (25) leads to:

$$\begin{aligned} y_N^{N,1} - y_N^{N-1,1} &= 0 \\ y_N^{N,2} - y_N^{N-1,2} &\approx B_N^1 \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) \\ y_N^{N,3} - y_N^{N-1,3} &\approx A_N^2 (y_N^{N,2} - y_N^{N-1,2}) + B_N^2 \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) \\ &= A_N^2 B_N^1 \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) + B_N^2 \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) \\ &= (A_N^2 B_N^1 + B_N^2) \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) \\ y_N^{N,m} - y_N^{N-1,m} &\approx \sum_{i=1}^{m-1} \left( B_N^i \prod_{j=i+1}^{m-1} A_N^j \right) \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}), \quad m \geq 2, \end{aligned} \quad (27)$$

where we used in the first equation that both Markov chains start from the same initial condition  $y_N^{N,1} = y_N^{N-1,1} = y_N^{N-1,M}$ . Substituting equation (27) into equation (22) we find:

$$f(\bar{X}^N, y_N^{N,m}) - f(\bar{X}^{N-1}, y_N^{N-1,m}) \approx \left( \partial_x f^m + \partial_y f^m \sum_{i=1}^{m-1} \left( B_N^i \prod_{j=i+1}^{m-1} A_N^j \right) \right) \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}),$$

from which we obtain:

$$\hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N) - \hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N) \approx \frac{\Delta t}{M} \sum_{m=1}^M \left( \partial_x f^m + \partial_y f^m \sum_{i=1}^{m-1} \left( B_N^i \prod_{j=i+1}^{m-1} A_N^j \right) \right) \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}). \quad (28)$$

Substituting (28) into equation (18), we find an approximation for the variance of the estimator in (20):

$$\text{Var} \left[ \bar{F}_M^{\text{HMM}}(\bar{X}^N) \right] \approx \text{Var} \left[ \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) + \frac{\Delta t}{M} \sum_{m=1}^M \left( \partial_x f^m + \partial_y f^m \sum_{i=1}^{m-1} \left( B_N^i \prod_{j=i+1}^{m-1} A_N^j \right) \right) \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) \right]. \quad (29)$$

From equation (29), we can draw a number of conclusions:

- (i) the variance of the estimator at time  $t^N$  grows only slightly with respect to the variance at time  $t^{N-1}$ ;
- (ii) even when the variance at time  $t^{N-1}$  is zero (which happens when it is computed via a deterministic reinitialization), the variance at time  $t^N$  will be nonzero, since the coefficients  $A_N^j$  and  $B_N^j$  depend on the Brownian paths, see equation (26);
- (iii) the variance of the estimator will be an increasing function of  $N$ , since a bit of variance is added on at every macroscopic time step.

This last observation is the reason we introduced a reinitialization procedure in section 3.2.



#### 4.1.2. The linear case

For the linear system (5), we show that the variance vanishes exactly when the initial estimator  $\bar{F}_M^{\text{HMM}}(X^0)$  is deterministic. In this case, we have:

$$\partial_x f^m = \lambda, \quad \partial_y f^m = p, \quad \partial_x g^m = q, \quad \partial_y g^m = -A, \quad \partial_x \beta^m = \partial_y \beta^m = 0.$$

Due to the linearity of the system, equation (27) (as well as all following equations) become exact. Moreover, because  $\partial_x \beta^m = \partial_y \beta^m = 0$ , the quantities  $A_N^m$  and  $B_N^m$ , defined in (26), become deterministic:

$$A_N^m = 1 - A \frac{\delta t}{\varepsilon}, \quad B_N^m = q \frac{\delta t}{\varepsilon}. \quad (30)$$

Thus, equation (27) can be rewritten as:

$$y_N^{N,m} - y_N^{N-1,m} = q \frac{\delta t}{\varepsilon} \left( \sum_{i=0}^{m-2} \left( 1 - A \frac{\delta t}{\varepsilon} \right)^i \right) \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}), \quad m \geq 2.$$

Equation (28) reads:

$$\hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N) - \hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N) = \left( \lambda + \frac{\delta t}{\varepsilon} \frac{pq}{M} \sum_{m=2}^M \left( \sum_{i=0}^{m-2} \left( 1 - A \frac{\delta t}{\varepsilon} \right)^i \right) \right) \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}), \quad (31)$$

which is deterministic as soon as  $\bar{F}_M^{\text{HMM}}(\bar{X}^{N-1})$  is deterministic. Using the known result on sums of geometric sequences, the sums in equation (31) can be further calculated as:

$$\sum_{m=2}^M \left( \sum_{i=0}^{m-2} (1 - \tilde{A})^i \right) = \frac{M}{\tilde{A}} (1 - \tilde{B}), \quad (32)$$

where we introduced the following two constants:

$$\tilde{A} = A \frac{\delta t}{\varepsilon}, \quad \tilde{B} = \frac{1 - (1 - \tilde{A})^M}{M \tilde{A}}. \quad (33)$$

Combining equations (31) and (32) and substituting the result into equation (39), we find the following expression for the variance-reduced estimator in the linear case:

$$\bar{F}_M^{\text{HMM}}(\bar{X}^N) = \left( 1 + \Delta t \left( \lambda + \frac{pq}{A} (1 - \tilde{B}) \right) \right) \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1}) \quad (34)$$

$$= \left( 1 + \Delta t \left( \lambda + \frac{pq}{A} (1 - \tilde{B}) \right) \right)^N \bar{F}_M^{\text{HMM}}(\bar{X}^0). \quad (35)$$

For the linear system, the expression of the variance in equation (29) is then given by:

$$\text{Var}[\bar{F}_M^{\text{HMM}}(\bar{X}^N)] = \left( 1 + \Delta t \left( \lambda + \frac{pq}{A} (1 - \tilde{B}) \right) \right)^{2N} \text{Var}[\bar{F}_M^{\text{HMM}}(\bar{X}^0)]. \quad (36)$$

Equation (36) reveals that in the linear case the variance of the proposed estimator at time step  $t^N$  depends only on the variance of the initial estimation. If a deterministic initialization is used in the first forward Euler step in equation (17), the variance vanishes since  $\text{Var}[\bar{F}_M^{\text{HMM}}(\bar{X}^0)] = 0$ .

This (perhaps) surprising result can also be seen as follows. In the second forward Euler step of the macroscopic equation, combining equation (17) with (18), the variance of the variance-reduced estimator is readily obtained as:

$$\begin{aligned} \text{Var}[\bar{F}_M^{\text{HMM}}(\bar{X}^1)] &= \text{Var} \left[ \hat{F}_M^{\text{HMM}}(\bar{X}^1; \omega_1) - \left( \hat{F}_M^{\text{HMM}}(\bar{X}^0; \omega_1) - F(\bar{X}^0) \right) \right] \\ &= \text{Var} \left[ \frac{1}{M} \sum_{m=1}^M \left( \lambda(\bar{X}^1 - \bar{X}^0) + p(y_1^{1,m} - y_1^{0,m}) \right) \right] \\ &= \frac{p^2}{M^2} \text{Var} \left[ \sum_{m=1}^M (y_1^{1,m} - y_1^{0,m}) \right], \end{aligned} \quad (37)$$

in which we used that  $\bar{X}^0$ ,  $\bar{X}^1$  and  $\bar{F}_M^{\text{HMM}}(\bar{X}^0)$  are all deterministic quantities. To calculate the sum in equation (37), we subtract the equations of the two Markov chains given in equation (14) from each other resulting in:

$$y_1^{1,m+1} - y_1^{0,m+1} = \left(1 - A \frac{\delta t}{\varepsilon}\right)(y_1^{1,m} - y_1^{0,m}) + \frac{\delta t}{\varepsilon}q(\bar{X}^1 - \bar{X}^0) + \sqrt{\frac{\delta t}{\varepsilon}}(\xi_1^m - \xi_1^m). \quad (38)$$

Since we are using the same Brownian path for both Markov chains, the stochastic part of equation (38) cancels out exactly. Therefore, this difference between Markov chain generated samples is completely deterministic and its variance is zero. This continues to hold for all following forward Euler steps.

## 4.2. Estimator bias

### 4.2.1. General case

Next, we examine the bias of the variance-reduced HMM estimator at time instant  $t^N$ , with respect to the original HMM estimator. Working out the recursion in equation (18), the estimator can be written as follows:

$$\begin{aligned} \bar{F}_M^{\text{HMM}}(\bar{X}^N) &= \hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N) - \left(\hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N) - \bar{F}_M^{\text{HMM}}(\bar{X}^{N-1})\right) \\ &= \hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N) + \left(\hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_{N-1}) - \hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N)\right) \\ &\quad - \left(\hat{F}_M^{\text{HMM}}(\bar{X}^{N-2}; \omega_{N-1}) - \bar{F}_M^{\text{HMM}}(\bar{X}^{N-2})\right) \\ &= \hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N) + \sum_{n=1}^{N-1} \left(\hat{F}_M^{\text{HMM}}(\bar{X}^n; \omega_n) - \hat{F}_M^{\text{HMM}}(\bar{X}^n; \omega_{n+1})\right) \\ &\quad - \left(\hat{F}_M^{\text{HMM}}(\bar{X}^0; \omega_1) - \bar{F}_M^{\text{HMM}}(\bar{X}^0)\right). \end{aligned} \quad (39)$$

$$\quad (40)$$

Taking the expectation of both sides of equation (40) over repeated experiments while keeping the sample size  $M$  and time step  $\delta t$  fixed leads to:

$$\mathbb{E}[\bar{F}_M^{\text{HMM}}(\bar{X}^N)] = \mathbb{E}[\hat{F}_M^{\text{HMM}}(\bar{X}^N; \cdot)] - \left(\mathbb{E}[\hat{F}_M^{\text{HMM}}(\bar{X}^0; \cdot)] - \mathbb{E}[\bar{F}_M^{\text{HMM}}(\bar{X}^0)]\right). \quad (41)$$

From equation (41), we observe that the variance-reduced HMM estimator  $\bar{F}_M^{\text{HMM}}(\bar{X}^N)$  does not introduce an additional bias compared to the original estimator  $\hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N)$ , provided that the initial variance-reduced estimator  $\bar{F}_M^{\text{HMM}}(\bar{X}^0)$  is unbiased with respect to  $\hat{F}_M^{\text{HMM}}(\bar{X}^0; \omega_0)$ . One way of ensuring this is to use a Metropolized version of the HMM estimators, such that each individual term in equation (41) is unbiased with respect to the exact time derivate  $F(\cdot)$ . (Note, however, that applying the Metropolis correction to both  $\hat{F}_M^{\text{HMM}}(\bar{X}^N; \omega_N)$  and  $\hat{F}_M^{\text{HMM}}(\bar{X}^{N-1}; \omega_N)$  may result in different samples getting rejected and hence a reduced correlation between the two estimator.) Alternatively, one could try to ensure that the initial variance-reduced estimator  $\bar{F}_M^{\text{HMM}}(\bar{X}^0)$  contains exactly the same bias as  $\hat{F}_M^{\text{HMM}}(\bar{X}^0; \omega_0)$ . These effects are illustrated numerically in section 5.1.

### 4.2.2. The linear case

As equation (41) shows, an additional bias may appear if the expectation of the initial estimator  $\bar{F}_M^{\text{HMM}}(\bar{X}^0)$  and of the standard HMM estimator  $\hat{F}_M^{\text{HMM}}(\bar{X}^0; \omega_0)$  are different. We now calculate the resulting bias on the solution paths obtained with the variance-reduced HMM technique for the linear system (5). We first write the forward Euler solution of the averaged equation (8) using the exact expression of  $F$  given in equation (7) as:

$$\begin{aligned} X^N &= X^{N-1} + \Delta t F(X^{N-1}) = \left(1 + \Delta t \left(\lambda + \frac{pq}{A}\right)\right) X^{N-1} \\ &= \left(1 + \Delta t \left(\lambda + \frac{pq}{A}\right)\right)^N X^0, \end{aligned} \quad (42)$$

where the last line is obtained by working out the recursion.

We want to write a similar expression for the variance-reduced solution paths. To that end, using equation (35), we write:

$$\begin{aligned}\bar{X}^N &= X^0 + \Delta t \sum_{n=0}^{N-1} \bar{F}_M^{\text{HMM}}(\bar{X}^n) \\ &= X^0 + \Delta t \bar{F}_M^{\text{HMM}}(\bar{X}^0) \sum_{n=0}^{N-1} \left(1 + \Delta t \left(\lambda + \frac{pq}{A}(1 - \tilde{B})\right)\right)^n,\end{aligned}\quad (43)$$

in which  $\tilde{B}$  is defined in (33) and the last line is again obtained by working out the recursion. The sum in equation (43) corresponds to a geometric sequence and can be calculated as:

$$\sum_{n=0}^{N-1} \left(1 + \Delta t \left(\lambda + \frac{pq}{A}(1 - \tilde{B})\right)\right)^n = \frac{\left(1 + \Delta t \left(\lambda + \frac{pq}{A}(1 - \tilde{B})\right)\right)^N - 1}{\Delta t \left(\lambda + \frac{pq}{A}(1 - \tilde{B})\right)}.\quad (44)$$

Substituting this into equation (43) yields:

$$\bar{X}^N = X^0 + \frac{\left(1 + \Delta t \left(\lambda + \frac{pq}{A}(1 - \tilde{B})\right)\right)^N - 1}{\left(\lambda + \frac{pq}{A}(1 - \tilde{B})\right)} \bar{F}_M^{\text{HMM}}(\bar{X}^0).\quad (45)$$

Equation (45) provides the variance-reduced solution path at time  $t^N$  which depends solely on the initial estimation  $\bar{F}_M^{\text{HMM}}(\bar{X}^0)$ . It is straightforward to verify that this formula delivers the expected results for  $N = 0$  and  $N = 1$ . In addition, equation (45) allows to calculate the asymptotic ( $N \rightarrow \infty$ ) behavior of  $\bar{X}^N$  for fixed  $\Delta t$ . Since  $M > 0$  and  $\tilde{A} = A \frac{\delta t}{\varepsilon} \leq 2$  due to stability of the Euler-Maruyama scheme in equation (14), we obtain:

$$1 - \tilde{B} \leq 1 < -\frac{\lambda A}{pq},\quad (46)$$

with  $\tilde{B}$  defined in (33), and hence equation (45) yields:

$$\lim_{N \rightarrow \infty} \bar{X}^N = \bar{X}^\infty = \bar{X}^0 - \frac{\bar{F}_M^{\text{HMM}}(\bar{X}^0)}{\lambda + \frac{pq}{A}(1 - \tilde{B})}.\quad (47)$$

When using an exact initialization  $\bar{F}_M^{\text{HMM}}(\bar{X}^0) = F(X^0)$ , we find:

$$\bar{X}^\infty = \bar{X}^0 - \frac{\left(\lambda + \frac{pq}{A}\right) \bar{X}^0}{\lambda + \frac{pq}{A}(1 - \tilde{B})} = -\frac{\frac{pq}{A} \tilde{B} \bar{X}^0}{\left(\lambda + \frac{pq}{A}(1 - \tilde{B})\right)}\quad (48)$$

From equation (48) we learn that the factor  $\tilde{B}$  determines the additional bias in solution paths of the linear system. To avoid a bias compared to the forward Euler solution with exact  $F$  we require that  $\tilde{B} = 0$ . Using the expression of  $\tilde{B}$  in equation (33) the following condition arises:

$$\tilde{B} = 0 \Rightarrow \left(1 - A \frac{\delta t}{\varepsilon}\right)^M = 1,$$

which is satisfied in the following three cases: (i)  $M = 0$ , (ii)  $\varepsilon$  constant and  $\delta t \rightarrow 0$ , and (iii) if the sample size  $M$  is even and at the same time the following constraint holds:

$$A \frac{\delta t}{\varepsilon} = 2.\quad (49)$$

In the numerical experiments in section 5, we only consider the third case.

## 5. Numerical results

We now put the variance reduction method to the test. We begin by looking at the bias and variance of the method after one iteration in section 5.1. We then turn to more detailed experiments on local variance reduction as a function of the numerical parameters in section 5.2 and on the resulting variance on the solution trajectories in 5.3. In all cases, we compare the linear and nonlinear model problems (5) and (10).

### 5.1. Bias and variance of a single estimation

Here, we investigate the effect of the three possible initialization procedures described in section 3.2. Since the first forward Euler step in any variance-reduced scheme needs to be taken with an accurate (low-variance) initial estimate  $\bar{F}_M^{\text{HMM}}(X^0)$ , we compare all estimations at time  $t^1 = \Delta t$ , given the macroscopic state  $\bar{X}^1$ , obtained as:

$$\bar{X}^1 = X^0 + \Delta t \bar{F}_M^{\text{HMM}}(X^0). \quad (50)$$

We perform  $J_r$  realizations, denoted by  $(\bar{X}_j^1)_{j=1}^{J_r}$ , using different random seeds. In section 5.1.1, we discuss the linear model problem (5). Afterwards, in section 5.1.2, we consider the nonlinear system given in (10). In both sections, we set  $\Delta t = 0.02$ ,  $J_r = 500$  and  $\varepsilon = 10^{-3}$ .

#### 5.1.1. Linear system

For the linear system (5), we choose  $\lambda = -10$ ,  $p = 4$ ,  $q = 0.5$  and  $A = 1.2$ . The initial conditions are chosen as  $X^0 = x_0 = 1$  and  $y_0 = 1$ . To show the variance reduction of the proposed method, we also perform a reference simulation using HMM without variance reduction. For this simulation, we choose the Markov chain time step  $\delta t = \varepsilon$ , while for the variance-reduced estimator, we fix  $\delta t$  as given in equation (49).

*HMM without variance reduction.* We begin by computing the HMM estimator distribution without variance reduction at time  $t^1$  by calculating 500 realizations  $\hat{F}_M^{\text{HMM}}(\hat{X}_j^1; \omega_{1,j})$  using  $M = 50$  samples and different seeds  $(\omega_{1,j})_{j=1}^{J_r}$ . Recall that the samples are generated by simulating the fast equation of system (5) using the Euler-Maruyama scheme given in equation (14), while keeping the current value of the slow variable fixed. The first forward Euler step is performed similar to (50):

$$\hat{X}_j^1 = X^0 + \Delta t \hat{F}_M^{\text{HMM}}(X^0; \omega_{0,j}), \quad 1 \leq j \leq J_r, \quad (51)$$

with different seeds  $(\omega_{0,j})_{j=1}^{J_r}$  in each realization. We compare the difference between estimations with and without the Metropolis-Hastings correction. In figure 2 (left), we plot the estimator distributions at time  $t^1$ , in which the blue and green dot-dashed lines represent the distributions without and with the Metropolis-Hastings algorithm, respectively. The vertical lines of the same color depict the sample mean of each distribution. The vertical red line corresponds to the exact mean, given by  $F(X^1)$ , with  $X^1$  the forward Euler solution of the averaged equation using the exact expression of  $F(x)$  in equation (7). We observe that, in the linear case, the HMM estimator is unbiased with respect to the forward Euler solution for  $M \rightarrow \infty$ . (This is due to the fact that the Markov chain generated by the Euler-Maruyama scheme (14) preserves the mean of the fast variable  $y$ , which is the only information on  $y$  that is used in the effective equation (8).) Moreover, the addition of the extra Metropolis step has no significant influence on the variance of the estimator  $\hat{F}_M^{\text{HMM}}(\bar{X}^1; \omega_1)$ .

*HMM with variance reduction without Metropolis-Hastings.* Next, we calculate the variance-reduced estimator distribution at time  $t^1$  and examine the effect of the different possible initializations (choices for  $\bar{F}_M^{\text{HMM}}(X^0)$ ). In the middle plot of figure 2, we show the results without adding the Metropolis-Hastings correction at  $t^0$ . The blue and green dot-dashed lines correspond to the distributions when using an estimated initialization  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_{0,j})$  and an averaged initialization  $\hat{F}_{10,50}^{\text{HMM}}(X^0; \omega_{0,j})$  in each realization  $j = 1, \dots, J_r$ , respectively. The pink line represents the estimator distribution when using the exact expression  $F(X^0)$  as initial estimator. Since the variance-reduced estimator yields a variance-free result when choosing an exact initialization in the linear case, see equation (36), the corresponding distribution reduces to a vertical pink

### Estimator distributions (linear case)

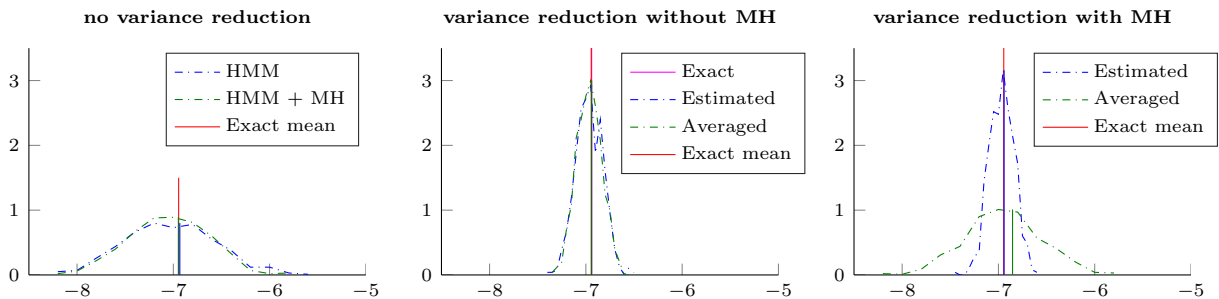


Figure 2: Estimator distributions after one time step  $\Delta t = 0.02$  when applied to the linear system (5). The exact mean  $F(\bar{X}^1)$  is depicted by a vertical red line in each plot. Left: HMM without variance reduction with (green) and without (blue) Metropolis-Hastings. Middle: variance-reduced HMM without Metropolis-Hastings for different initializations. Right: variance-reduced HMM with Metropolis-Hastings used in the estimated and averaged initializations.

line. The vertical red line corresponds to the exact mean  $F(X^1)$ . Since all mean values lie close together, the pink and red lines are hard to discern. We find that both initializations lead to an unbiased variance-reduced estimator and the reduction in variance is clearly visible. In figure 3, we more closely inspect the estimated and averaged initializations. On the left, we show the estimator distributions when using  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_{0,j})$  (solid blue) and  $\hat{F}_{5000}^{\text{HMM}}(X^0; \omega_{0,j})$  (dashed blue) as initial estimators. This confirms that the estimated initialization indeed leads to an unbiased estimator for  $M^* \rightarrow \infty$  and yields reductions in variance by a factor 12 and 118, respectively. On the right, we visualize the distributions when using  $\hat{F}_{10,50}^{\text{HMM}}(X^0; \omega_{0,j})$  (solid green),  $\hat{F}_{100,50}^{\text{HMM}}(X^0; \omega_{0,j})$  (dashed green) and  $\hat{F}_{1000,50}^{\text{HMM}}(X^0; \omega_{0,j})$  (dot-dashed green). This shows that the averaged initialization also gives rise to an unbiased estimator for  $S \rightarrow \infty$  with  $M$  fixed and yields reductions by a factor 13, 115 and 1163, respectively.

*HMM with variance reduction with Metropolis-Hastings.* When using Metropolis-Hastings to generate the samples in the initialization  $\bar{F}_M^{\text{HMM}}(X^0)$ , we obtain the distributions in figure 2 (right). The effect of using  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_{0,j})$  and  $\hat{F}_{10,50}^{\text{HMM}}(X^0; \omega_{0,j})$  as initial estimators is shown by blue and green dot-dashed lines, respectively. The former leads to an unbiased estimator with clear variance reduction, while the latter results in a biased estimator with only little reduction in variance. We regard both initializations in more detail in figure 4. By comparing the effect of using  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_{0,j})$  (blue solid) and  $\hat{F}_{5000}^{\text{HMM}}(X^0; \omega_{0,j})$  (blue dashed) in the left plot, we conclude that the estimated initialization yields an unbiased estimator for  $M^* \rightarrow \infty$  with reduction factors 11 and 106. The right plot indicates that by using Metropolis-Hastings in the averaged initialization the resulting estimator becomes unbiased for  $S \rightarrow \infty$  with  $M$  fixed. Moreover, the reduction in variance is significantly lower with corresponding reduction factors 1, 2 and 21. Although both initializations  $\hat{F}_{M^*}^{\text{HMM}}(X^0; \omega_{0,j})$  and  $\hat{F}_{M,S}^{\text{HMM}}(X^0; \omega_{0,j})$  are computationally equivalent when  $M^* = SM$ , the former outperforms the latter by far, which requires further research to better understand this behavior.

#### 5.1.2. Nonlinear system

We now consider the nonlinear system (10), and set the initial conditions as  $x_0 = 1$  and  $y_0 = 0.5$ . Furthermore, we set the time step  $\delta t = \varepsilon$ . In the linear case, we compared our results with the exact mean at time  $t^1$ , given by  $F(X^1)$ . However, since the HMM estimator is biased in the general nonlinear case and we are only interested in studying the bias in the estimation (and not in the solution paths  $\bar{X}^1$ ), here, we define the exact mean for the HMM estimator without and with variance reduction as:

$$\hat{m}_e = \frac{1}{J_r} \sum_{j=1}^{J_r} F(\hat{X}_j^1), \quad \bar{m}_e = \frac{1}{J_r} \sum_{j=1}^{J_r} F(\bar{X}_j^1), \quad (52)$$

Variance-reduced estimator distributions without MH (linear case)

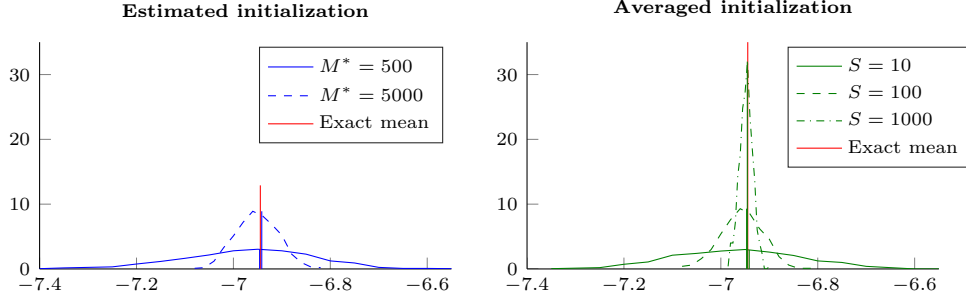


Figure 3: Variance-reduced estimator distributions (without Metropolis-Hastings) after one time step  $\Delta t = 0.02$  when applied to the linear system (5). The exact mean  $F(X^1)$  is depicted by a vertical red line in each plot. Left: estimated initialization. Right: averaged initialization.

Variance-reduced estimator distributions with MH (linear case)

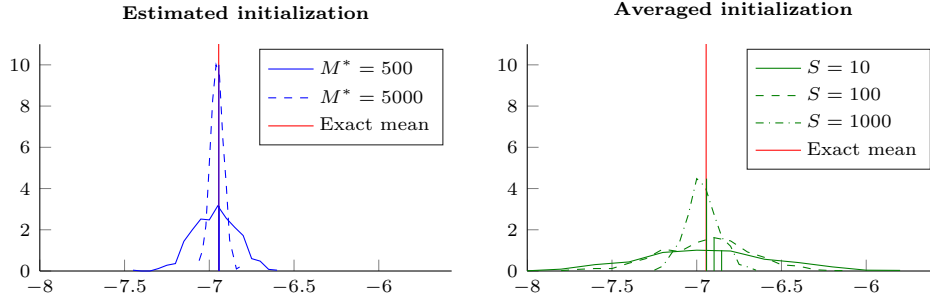


Figure 4: Variance-reduced estimator distributions (including Metropolis-Hastings) after one time step  $\Delta t = 0.02$  when applied to the linear system (5). The exact mean  $F(X^1)$  is depicted by a vertical red line in each plot. Left: estimated initialization. Right: averaged initialization.

in which  $F(x)$  corresponds to the exact right hand side of the averaged equation given in equation (11).

*HMM without variance reduction.* When computing the HMM estimator distribution at time  $t^1$  without and with Metropolis-Hastings, we obtain the blue and green dot-dashed distributions in figure 5 (left), respectively. The vertical red lines correspond to the exact means  $\hat{m}_e$  given in equation (52) for both experiments, which coincide to the naked eye. We immediately see the necessity of including the Metropolis-Hastings algorithm: the blue distribution has a very clear bias after only one iteration of the method, while the green distribution possesses the correct mean.

*HMM with variance reduction without Metropolis-Hastings.* Repeating the above experiment for the variance-reduced estimator without Metropolis-Hastings at time  $t^1$  using different initializations, we show the results in the middle plot of figure 5. The blue and green dot-dashed lines represent the distributions when using  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_{0,j})$  and an  $\hat{F}_{10,50}^{\text{HMM}}(X^0; \omega_{0,j})$  in each realization  $j = 1, \dots, J_r$ , respectively. When using an exact initialization  $F(X^0)$ , we obtain the sharply peaked pink distribution centered around the exact means  $\bar{m}_e$  (vertical red lines) in equation (52) of these three experiments, which again coincide to the naked eye. From this, we find that, while the variance is significantly reduced (reduction factors 11, 12 and 1219), the variance-reduced estimator is biased for the estimated and averaged initializations. Moreover, the resulting estimator is only unbiased when using an exact initialization, leading to an even stronger reduction in variance.

Estimator distributions (nonlinear case)

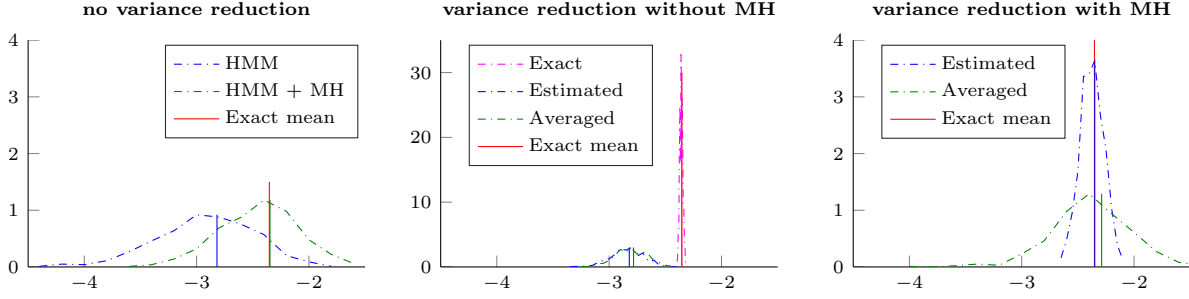


Figure 5: Estimator distributions after one time step  $\Delta t = 0.02$  when applied to the nonlinear system (10). The exact mean given in (52) is depicted by a vertical red line in each plot. Left: HMM without variance reduction with (green) and without (blue) Metropolis-Hastings. Middle: variance-reduced HMM without Metropolis-Hastings for different initializations. Right: variance-reduced HMM with Metropolis-Hastings used in the estimated and averaged initializations.

*HMM with variance reduction with Metropolis-Hastings.* Since the exact initialization used in the previous experiment is generally not possible, we apply the Metropolis correction in the initialization to avoid the bias. The results are shown in figure 5 (right), in which the blue and green dot-dashed lines correspond to estimator distributions using an estimated and averaged initialization, respectively. In this case, we derive the same conclusion as for the linear system (see figure 2, right): the blue distribution is unbiased and shows a clear reduction in variance, while the green distribution is biased and gives only little reduction. To conclude, we investigate the estimated and averaged initializations with Metropolis-Hastings in figure 6. It is seen that using  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_{0,j})$  and  $\hat{F}_{5000}^{\text{HMM}}(X^0; \omega_{0,j})$  lead to an unbiased variance-reduced estimator with reduction factors 10 and 104 (left plot), while using  $\hat{F}_{S,M}^{\text{HMM}}(X^0; \omega_{0,j})$  becomes unbiased for fixed  $M$  and  $S \rightarrow \infty$  with reduction factors 1, 2 and 22 (right plot).

### 5.2. Local variance reduction

In this section, we demonstrate the local reduction in variance of the different estimators by considering their variance after  $N = 1$ ,  $N = 4$  and  $N = 10$  time steps of the method. The variance is obtained by repeating the computations over 100 realizations. For a fixed number of time steps, we study the influence of the macroscopic time step  $\Delta t$  by varying it as:

$$\Delta t = [0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001]. \quad (53)$$

We apply the estimators both to the linear and nonlinear system using the same parameters as in sections 5.1.1 and 5.1.2, respectively.

For the linear system (5), the results are plotted in figure 7. In each plot, the HMM estimator variance without variance reduction (blue dot-dashed line) is seen to be constant as a function of  $\Delta t$  and is unchanged when increasing the number of macroscopic time steps  $N$ . This is because As sthe variance of a Markov chain Monte Carlo method is essentially of the order  $1/M$  (here,  $M = 50$ ), thus constant in  $\Delta t$  and independent of time. On the contrary, the variance-reduced estimator with  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_0)$  (solid blue) and  $\hat{F}_{10,50}^{\text{HMM}}(X^0; \omega_0)$  (solid green) as initial estimators yields a variance that clearly depends on  $\Delta t$  and  $N$ . As indicated on figure 2 (middle), both initializations lead to variances that are much alike. We observe that, for fixed  $N$ , the variance becomes smaller for increasing time step  $\Delta t$ , while, for a fixed time step  $\Delta t$ , the variance decays rapidly with increasing  $N$ . This behavior is confirmed by our analysis, see equation (36), for which we have  $\tilde{B} = 0$  due to our choice of  $\delta t$  in equation (49). In that case, since  $\lambda + pq/A < 0$  due to stability, equation (36) shows that for fixed  $N$  the variance  $\text{Var}[\hat{F}_M^{\text{HMM}}(\bar{X}^N)]$  converges to  $\text{Var}[\hat{F}_M^{\text{HMM}}(\bar{X}^0)]$  from below for

Variance-reduced estimator distributions with MH (nonlinear case)

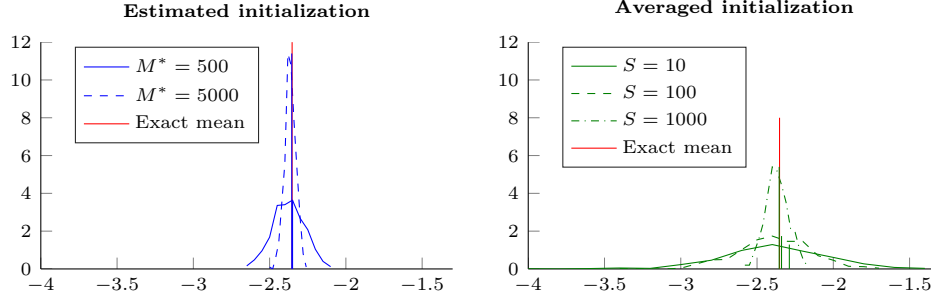


Figure 6: Variance-reduced estimator distributions (including Metropolis-Hastings) after one time step  $\Delta t = 0.02$  when applied to the nonlinear system (10). The exact mean given in (52) is depicted by a vertical red line in each plot. Left: estimated initialization. Right: averaged initialization.

$\Delta t \rightarrow 0$ . It also states that for fixed  $\Delta t$  the variance  $\text{Var}[\bar{F}_M^{\text{HMM}}(\bar{X}^N)]$  decays exponentially with increasing  $N$ . The expected evolution of variance in equation (36) is depicted by a dashed red line in each plot.

When repeating the above experiment for the nonlinear system (10), we obtain the plots in figure 8. Once more, the HMM estimator variance (blue dot-dashed line) is roughly constant in  $\Delta t$  and  $N$ . The red, blue and green solid lines correspond to the variance of the variance-reduced estimator when using  $F(X^0)$ ,  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_0)$  and  $\hat{F}_{10,50}^{\text{HMM}}(X^0; \omega_0)$  as initialization, respectively. The red line represents the best possible performance of the proposed variance reduction method. It confirms the formal result obtained in equation (29), that is: (i) the variance increases with increasing  $N$ ; (ii) although  $\text{Var}[\bar{F}_M^{\text{HMM}}(X^0)] = 0$  there is a small contribution to the variance due to the dependence of  $A_N^j$  and  $B_N^j$  on the Brownian path which is observed for  $N = 1$ ; (iii) since a little bit of variance is added in every macroscopic step, for  $N = 1$  the variance decays as  $\Delta t^2$  and this slope gradually decreases for increasing  $N$ . The solid blue and green line show that the estimated and averaged initializations give rise to a constant variance in  $\Delta t$  and  $N$  corresponding to the variance of the initial estimator. As observed in figure 8 (right), the averaged initialization resulted in practically no reduction in variance, which is clearly visible in each plot of figure 8.

### 5.3. Solution trajectories

Finally, we look at the solution paths of the averaged equation (2) obtained by different estimators for  $F(X)$ . We apply the method both to the linear (section 5.3.1) and nonlinear (section 5.3.2) system.

#### 5.3.1. Linear system

Here, we focus on approximating the reduced evolution of the slow variable in equation (8), which is, in turn, an approximation of the slow variable's true evolution described in system (5). We compute the solution for  $t \in [0, 1]$  using initial conditions  $X^0 = x_0 = 1$  and  $y_0 = 1$ . The system parameters in equation (5) are as follows:  $\lambda = -10$ ,  $p = 4$ ,  $q = 0.5$  and  $A = 1.2$ .

We begin by applying the HMM technique without variance reduction generating  $M = 50$  samples in each iteration. For stability, the time step  $\delta t$  used in the Euler-Maruyama discretization of the fast equation is chosen as  $\delta t = \varepsilon$  with  $\varepsilon = 10^{-3}$ . The forward Euler time step used in the discretization of the macroscopic equation (2) is fixed as  $\Delta t = 0.02$ . The time evolution of the variables  $X$  and  $F$  is depicted by the blue line in the left plots of figure 9. The red line represents the exact solution of the macroscopic equation given in equation (9). Clearly, the statistical error dominates, thus justifying the need for variance reduction. The variance on  $X$  and  $F$  can be seen by the blue line in the right plots of figure 9 and is calculated by repeating the above experiment 100 times. We observe that the variance of the HMM estimator remains constant in time and behaves as  $O(1/M)$  which is typical for a Markov chain Monte Carlo estimator.

Next, we examine the proposed variance reduction technique based on control variables, as introduced in section 3.2, for which we first need to specify the initial estimation  $\bar{F}_M^{\text{HMM}}(X^0)$ . As pointed out at the end



Local variance reduction (linear case)

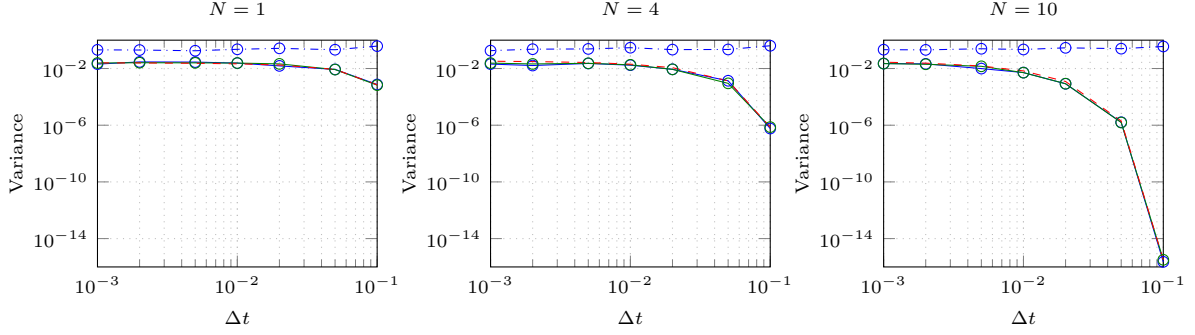


Figure 7: Local variance reduction of estimators evaluated after one (left), four (middle) and ten (right) time steps as a function of the macroscopic time step  $\Delta t$  for linear system (5). Blue dot-dashed line: HMM estimator without variance reduction; solid blue and green lines: variance-reduced HMM estimator using an estimated and averaged initialization, respectively; red dashed line: expected variance according to (36).

Local variance reduction (nonlinear case)

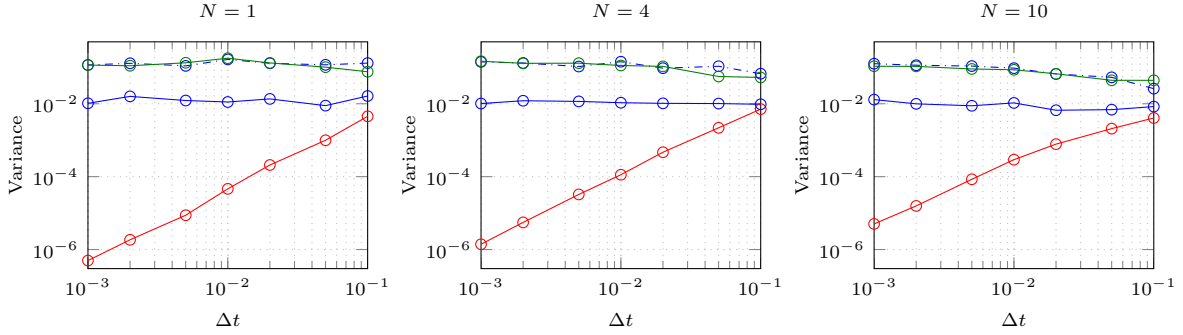


Figure 8: Local variance reduction of estimators evaluated after one (left), four (middle) and ten (right) time steps as a function of the macroscopic time step  $\Delta t$  for nonlinear system (10). Blue dot-dashed line: HMM estimator without variance reduction; solid red, blue and green lines: variance-reduced HMM estimator using an exact, estimated and averaged initialization, respectively.

of section 4.1 and observed in section 5.1.1, when using an exact initialization  $\bar{F}_M^{\text{HMM}}(X^0) = F(X^0)$  with  $F(x)$  calculated in equation (8), the variance-reduced estimator is completely variance-free, thus leading to a deterministic estimator. This is confirmed by the right plots of figure 9, which demonstrate that the variance on both  $X$  and  $F$  is indeed zero up to machine precision. Subsequently, we compare the effect when using an estimated initialization  $\hat{F}_{M^*}^{\text{HMM}}(X^0; \omega_0)$ . We remark that, to avoid introducing a bias in the linear case, the initial estimator is required to use a time step  $\delta t$  in the Euler-Maruyama scheme as derived in equation (49). In all plots of figure 9, the green line represents simulations when using  $M^* = 500$  samples in the initial estimator. The left plots show that both  $X$  and  $F$  evolve much smoother than its HMM counterpart. The bottom right plot confirms that the variance on  $F$  decays exponentially with time starting from the variance of the initial estimation, which was derived in equation (36). However, the top right plot indicates that there is no reduction in variance on the trajectories  $X$  with this initialization. To that end, when choosing  $M^* = 5000$  samples in the initial estimator, we can improve the reduction in variance for both  $X$  and  $F$  by a factor 10, which is shown by the cyan line on both right plots.

Solution trajectories and variance (linear case)

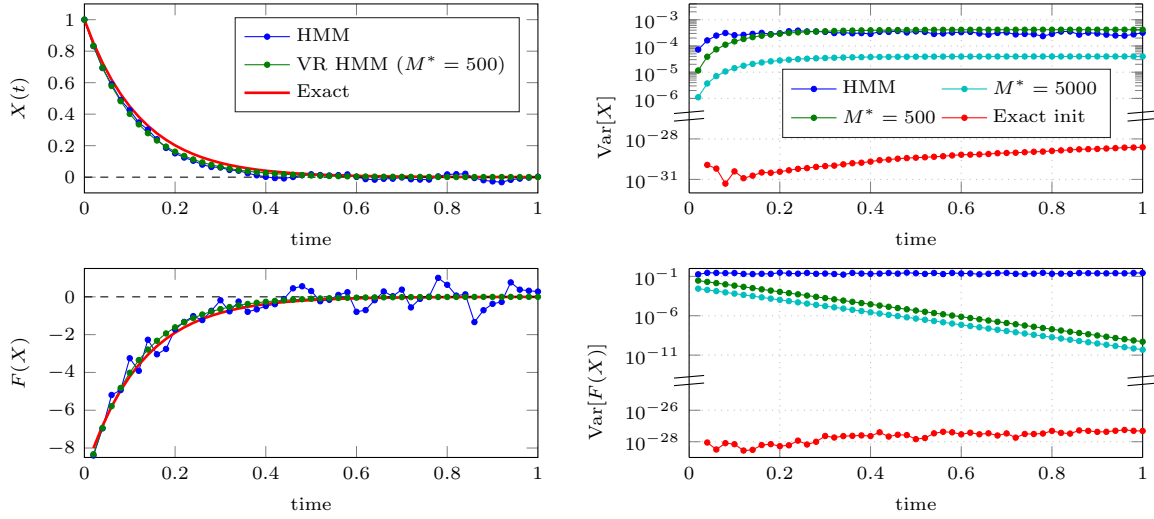


Figure 9: Left: time evolution of  $X$  and  $F$  when applying HMM to the linear model problem (5) with and without variance reduction (green and blue lines, respectively). The red line represents the exact solution. Right: variance on  $X$  and  $F$ . Blue line: HMM without variance reduction; red line: variance-reduced HMM with exact initialization; green and cyan lines: variance-reduced HMM with estimated initialization using  $M^* = 500$  and  $M^* = 5000$  samples, respectively.

### 5.3.2. Nonlinear system

As a second model problem, we consider the nonlinear stochastic multiscale system given in equation (10). We calculate the solution of the averaged equation for  $t \in [0, 2]$  using initial conditions  $X^0 = x_0 = 0.5$  and  $y_0 = 0.5$ .

We begin by applying the HMM procedure without variance reduction using  $M = 50$  samples by iterating over the Euler-Maruyama scheme (14) for the fast dynamics of system (10) with time step  $\delta t = \varepsilon$  and  $\varepsilon = 10^{-3}$ . We recall from section 5.1.2 that the HMM solution converges to the wrong solution (that is, the solution of the wrong equation) in the nonlinear case, due to the time discretization error of the Euler-Maruyama scheme. Therefore, we require the Metropolis correction to avoid this bias. The forward Euler time step used in the discretization of the macroscopic equation is fixed as  $\Delta t = 0.05$ . The time evolution of the variables  $X$  and  $F$  and their corresponding variance is depicted by the blue lines in figure 10. The red line in the left plots represents the exact solution for  $X$  and  $F$  as given in (12) and (11), respectively.

Next, we apply the variance-reduced HMM estimator. As noted in section 4.2.1, when combining the Metropolis-Hastings algorithm with the variance-reduced estimator, we lose strong correlation between the HMM estimators  $\hat{F}_M^{\text{HMM}}(\hat{X}^N; \omega_N)$  and  $\hat{F}_M^{\text{HMM}}(\hat{X}^{N-1}; \omega_N)$  in equation (18). This is clarified as follows. Both HMM estimators generate an ensemble of samples using the same seed  $\omega_n$  but a different value of the slow variable. In general, it is not known a priori when and which samples will be accepted or rejected in the Metropolis-Hastings algorithm. Since samples can be rejected in different places in both ensembles, they are in principle no longer correlated. Consequently, when subtracting these two estimators the statistical error will be larger than that of the individual estimators. To resolve this, we instead use two classical HMM estimators without the Metropolis-Hastings extension in equation (18), each producing a bias. However, since both contain the same bias, subtraction yields a result of order  $\Delta t$  (that is, the distance between  $\bar{X}^{N-1}$  and  $\bar{X}^N$ ) which lies within the accuracy of the forward Euler method.

In the variance-reduced setting, we perform experiments using both an exact initialization  $F(X^0)$ , as given in equation (11), as well as an estimated initialization  $\hat{F}_{M^*}^{\text{HMM}}(X^0; \omega_0)$  in the first forward Euler step of the averaged equation. The results are shown in figure 10, where the green lines depict the evolution of

Solution trajectories and variance (nonlinear case)

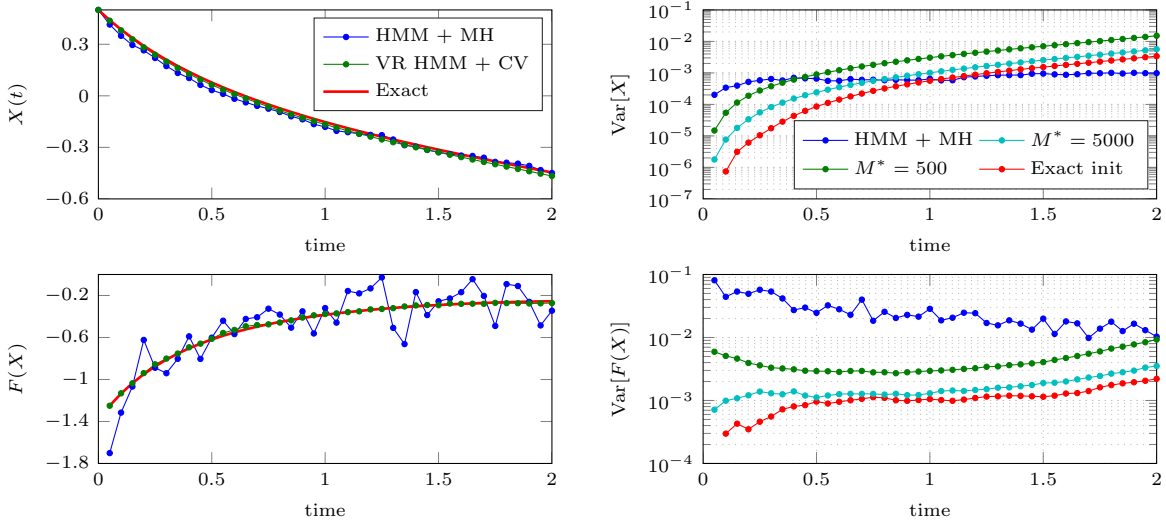


Figure 10: Left: time evolution of  $X$  and  $F$  when applying HMM with Metropolis-Hastings to the nonlinear model problem (10) with and without variance reduction (green and blue lines, respectively). The red line represents the exact solution. Right: variance on  $X$  and  $F$ . Blue line: HMM without variance reduction; red line: variance-reduced HMM with exact initialization; green and cyan lines: variance-reduced HMM with estimated initialization using  $M^* = 500$  and  $M^* = 5000$  samples, respectively.

the variables  $X$  and  $F$  and their corresponding variance when using  $M^* = 500$  in the initial estimator. In addition, in the right plots we also show the variance on  $X$  and  $F$  when choosing an exact initialization (red lines) and when setting  $M^* = 5000$  (cyan lines). From this, we observe that there is a clear buildup of variance for all variance-reduced estimators.

As suggested at the end of section 3.2, the variance buildup can be countered by occasionally reinitializing the estimator. Therefore, in what follows, we consider the influence of reinitializing the estimator after every  $R$  macroscopic time steps; that is, we compute a new accurate estimation (similar to the initialization) after a fixed number of macroscopic time steps using  $M_r$  samples. The resulting variance on  $X$  and  $F$  is plotted in figure 11, where we compare the HMM estimator using  $M = 50$  samples (solid blue line) with the variance-reduced estimator using  $\hat{F}_{500}^{\text{HMM}}(X^0; \omega_0)$  as initial estimator,  $M = 20$  samples in the HMM estimator difference in (18) for  $R = 20$  (green line),  $R = 10$  (red line),  $R = 5$  (cyan line) and  $R = 2$  (purple line) based on  $M_r = 500$  samples. It is seen that the variance on the solution trajectories clearly depends on the value of  $R$ . This experiment shows that, by repeatedly reinitializing the estimator, we can control the variance on both  $X$  and  $F$  and counter the buildup of variance on both quantities that was seen in figure 10.

## 6. Conclusions

We presented a variance reduction technique based on control variables for stochastic slow-fast systems containing a deterministic slow equation and a stochastic fast equation. The proposed method supplements the HMM estimator, which approximates the right hand side integral in the reduced description of the stochastic system, and we commented on its applicability for the coarse-projective integration estimator in the equation-free framework. We discussed the initialization of the variance-reduced estimator by means of an exact, estimated or averaged initial estimator and considered its reinitialization to counter the buildup of variance with time. We analyzed the estimator variance and additional bias compared to the HMM estimator and derived explicit expressions for a linear stochastic system. We applied the variance-reduced estimator to a linear and nonlinear test problem, in which we considered the effects of the initialization and

Solution trajectories and variance with reinitialization (nonlinear case)

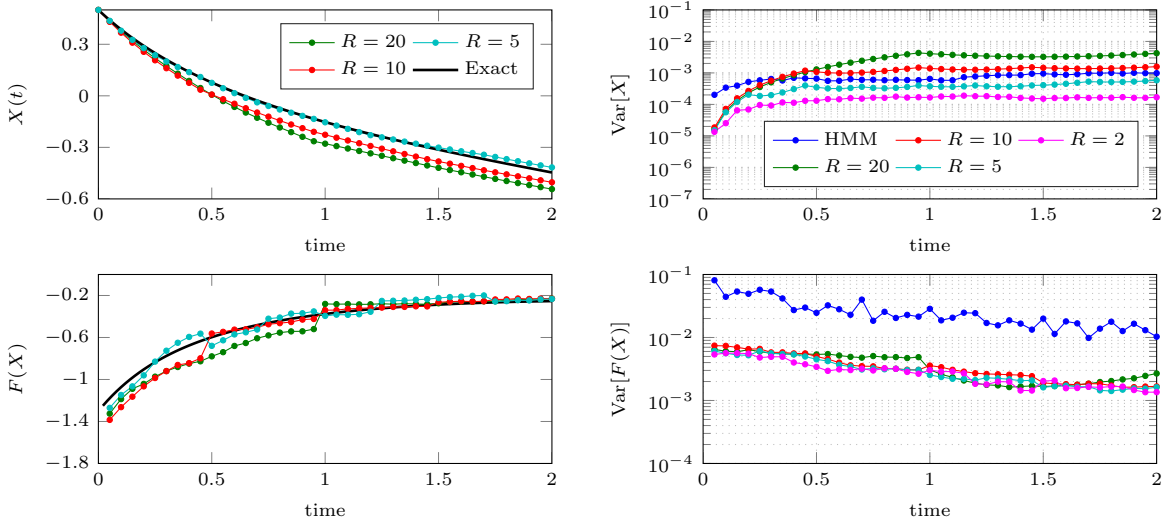


Figure 11: Left: time evolution of  $X$  and  $F$  when applying HMM with variance reduction to the nonlinear model problem (10) for different reinitialization frequencies  $R$ . The black line represents the exact solution. Right: variance on  $X$  and  $F$ . Blue line: HMM without variance reduction; other lines: variance-reduced HMM using an estimated initialization with  $M^* = 500$  and reinitializing after every  $R$  time steps, for  $R = 20$  (green line),  $R = 10$  (red line),  $R = 5$  (cyan line) and  $R = 2$  (purple line). The reinitialization uses the same procedure as the initialization.

reinitialization and compared the results with the classical HMM estimator. The numerical experiments showed that nonlinear systems generally require the Metropolis-Hastings correction in both estimators to avoid introducing a bias. Moreover, reinitializing the estimator was found to be very effective to control the variance on solution trajectories.

In the experiments, we did not compare in detail the computational cost of the different procedures for a desired variance. The variance-reduced scheme has many numerical parameters that can be chosen (such as the frequency and accuracy of reinitialization). Moreover, the standard HMM scheme benefits from a self-averaging effect, since the errors in subsequent macroscopic time steps are independent. In the variance-reduced scheme, the individual errors are smaller, but they are correlated. As a consequence, a fair comparison of computational cost is highly non-trivial and may well be problem-dependent. We postpone such a comparison to future research.

## References

## References

- [1] N. Berglund, B. Gentz, Noise-induced phenomena in slow-fast dynamical systems: a sample-paths approach, Springer Science & Business Media, 2006.
- [2] M. Bruna, S. J. Chapman, M. J. Smith, Model reduction for slow-fast stochastic systems with metastable behaviour, *Journal of Chemical Physics* 140 (17) (2014) 1–23.
- [3] P. Imkeller, J.-S. von Storch, Stochastic Climate Models, Birkhäuser Basel, Basel, 2001.
- [4] R. Erban, S. J. Chapman, I. G. Kevrekidis, T. Vejchodský, Analysis of a stochastic chemical system close to a sniper bifurcation of its mean field model, *SIAM Journal on Applied Mathematics* 70 (3) (2009) 984–1016.

- [5] D. Givon, R. Kupferman, A. Stuart, Extracting macroscopic dynamics: model problems and algorithms, *Nonlinearity* 17 (6) (2004) 55–127.
- [6] M. Rousset, G. Samaey, Individual-Based Models for Bacterial Chemotaxis in the Diffusion Asymptotics, *Mathematical Models and Methods in Applied Sciences* 23 (11) (2013) 2005–2037.
- [7] G. A. Pavliotis, A. Stuart, *Multiscale methods: averaging and homogenization*, Springer Science & Business Media, 2008.
- [8] T. Li, A. Abdulle, W. E, Effectiveness of implicit methods for stiff stochastic differential equations, *Communications in Computational Physics* 3 (2) (2008) 295–307.
- [9] E. Vanden-Eijnden, Numerical techniques for multi-scale dynamical systems with stochastic effects, *Communications in Mathematical Sciences* 1 (2) (2003) 385–391.
- [10] W. E, D. Liu, E. Vanden-Eijnden, Analysis of Multiscale Methods for Stochastic Differential Equations, *Communications on Pure and Applied Mathematics* 58 (11) (2005) 1544–1585.
- [11] R. E. Caflisch, Monte carlo and quasi-monte carlo methods, *Acta numerica* 7 (1998) 1–49.
- [12] W. E, B. Engquist, The Heterogeneous Multiscale Methods, *Communications in Mathematical Sciences* 1 (1) (2003) 87–132.
- [13] A. Abdulle, W. E, B. Engquist, E. Vanden-Eijnden, The heterogeneous multiscale method, *Acta Numerica* 21 (May 2012) (2012) 1–87.
- [14] D. Givon, I. G. Kevrekidis, R. Kupferman, Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems, *Communications in Mathematical Sciences* 4 (4) (2006) 707–729.
- [15] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, C. Theodoropoulos, Equation-Free, Coarse-Grained Multiscale Computation: enabling microscopic simulators to perform system-level tasks, *Communications in Mathematical Sciences* 1 (4) (2003) 715–762.
- [16] I. G. Kevrekidis, G. Samaey, Equation-free multiscale computation: algorithms and applications., *Annual review of physical chemistry* 60 (2009) 321–344.
- [17] C. W. Gear, I. G. Kevrekidis, C. Theodoropoulos, ‘Coarse’ integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods, *Computers and Chemical Engineering* 26 (7-8) (2002) 941–963.
- [18] W. Melis, G. Samaey, Variance-reduced HMM for stochastic slow-fast systems, *Procedia Computer Science* 80 (2016) 1255–1266.
- [19] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer Science & Business Media, 2003.
- [20] A. Papavasiliou, I. G. Kevrekidis, Variance reduction for the equation-free simulation of multiscale stochastic systems, *Multiscale Modeling & Simulation* 6 (1) (2007) 70–89.
- [21] D. J. Higham, An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, *SIAM Review* 43 (3) (2001) 525–546.
- [22] G. O. Roberts, R. L. Tweedie, Exponential convergence of langevin distributions and their discrete approximations, *Bernoulli* (1996) 341–363.
- [23] E. Cancès, F. Legoll, G. Stoltz, Theoretical and numerical comparison of some sampling methods for molecular dynamics, *ESAIM: M2AN* 41 (2) (2007) 351–389.
- [24] I. Fatkullin, E. Vanden-Eijnden, A computational strategy for multiscale systems with applications to lorenz 96 model, *Journal of Computational Physics* 200 (2) (2004) 605–638.